DOCUMENT RESUME

ED 426 088 TM 029 316

AUTHOR Tanguma, Jesus
TITLE Analyzing Repeated Measures Designs Using Univariate and Multivariate Methods: A Primer.
PUB DATE 1999-01-00
NOTE 35p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 21-23, 1999).
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Algebra; *Multivariate Analysis; Performance Factors; *Research Design; Research Methodology; Tables (Data)
IDENTIFIERS Homogeneity of Variance; *Repeated Measures Design; Sphericity Tests; Type I Errors; Type II Errors; *Univariate Analysis

ABSTRACT

Similarities and differences in the univariate and multivariate analysis of repeated measures designs are discussed, using a hypothetical data set studying the effects of practice on the algebra performance of four students to illustrate both methods. When data are analyzed through the univariate approach and the homogeneity assumption is violated, three correcting factors are presented. When data are analyzed using the multivariate approach, the homogeneity assumption is not necessary. The paper also presents the effects on Type I and Type II error rates of violating or not violating the assumption of homogeneity of variance. Each approach has its own assumptions to meet, but the sphericity assumption of the univariate approach is almost always violated. Even when the normality assumption of the multivariate approach is violated, such violations are generally regarded as less serious than violations of the sphericity assumption. When the researcher's concern is committing a Type I or Type II error, and several assumptions hold, the multivariate approach is suggested. (Contains 11 tables, 2 figures, and 11 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 088

Running head: REPEATED MEASURES

Analyzing Repeated Measures Designs Using Univariate and Multivariate Methods: A Primer

Jesus Tanguma

Texas A & M University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Jesus Tanguma

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

TM029316

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, January 1999.

Abstract

The present paper presents similarities and differences between the univariate and the multivariate analysis of repeated measures designs. Both methods are illustrated by means of an example. When the data are analyzed using the univariate approach and the homogeneity assumption is violated, three correcting factors are presented. When the data are analyzed using the multivariate approach, the homogeneity assumption is not necessary. The paper also presents the effects on the Type I and/or Type II error rates of violating or not violating the assumption of homogeneity of variance.

Analyzing Repeated Measures Designs Using Univariate and Multivariate Methods: A Primer

Researchers, in an effort to reduce error variance and systematic bias, typically assign subjects randomly to the different treatments in the experiment (Stevens, 1996). If each subject in the experiment is given only one treatment, the design is called a completely randomized block design (also known as a between-subjects design). However, if each subject is given two or more treatments, the design is called a repeated measures design (also known as a within-subjects design). Since repeated measures designs involve each subject being measured more than once on the same variable, such designs require less subjects for a given study.

For example, suppose a researcher is investigating the effect of three different sleeping aid pills. A between-subjects design would require three different groups of individuals. Consequently, if, say, each group were to have 5 subjects in it, analyzing the data using a between-subjects design would require 15 subjects. However, if the same individual is allowed to participate in all the conditions of the study (i.e., the data are created and analyzed using a within-subjects design), only 5 individuals would be required. Thus, it follows that when subjects are scarce or observations are expensive to obtain, repeated measures designs are more economical than a corresponding between-subjects design.

The purpose of the present paper is to discuss the similarities and differences between the univariate and the multivariate analysis of repeated measures designs. To do so, a hypothetical data set will be presented and analyzed using both methods.

Advantages and Disadvantages

Advantages of Repeated Measures Designs

As stated by Keppel and Saufley (1980), " The within-subjects design has become the typical design used to study such phenomena as learning, transfer of training, and practice effects of all sorts" (p. 175). In a pretest-posttest design, for example, subjects are observed at pretest, receive a treatment, and are then observed at posttest. Thus, the researcher has two observations per subject in the study. However, if a retention test is administered at a later date (e.g., one week later), then the researcher has three observations on each subject. Another example might be when police officer trainees are learning how to properly handcuff an individual. In this situation, the trainee is allowed to perform the particular task several times. After each practice trial, the trainee's performance is assessed. The researcher can then determine how trainees improve over repeated trials.

In addition to being economical as regards the number of subjects required for a given experiment, Neter, Kutner, Nachtsheim, and Wassarman (1996) note that

> A principal advantage of repeated measures designs is that they provide good precision for comparing treatments because all sources of variability between subjects are excluded from the experimental error. Only variation within the subjects enters the experimental error, since any two treatments can be compared directly for each subject. Thus, one may view the subjects as serving as their own controls. (p. 1165)

Since all variability due to individual differences has been excluded from the experimental error term, repeated measures designs are "much more powerful than completely randomized designs, where different subjects are randomly assigned to the different treatments" (Stevens, 1996, p. 450). Another advantage of the within-subjects designs is that, since the same subjects are being observed repeatedly, the researcher does not have to repeat the instructions.

Disadvantages of Repeated Measures Designs

Repeated measures designs have several disadvantages, "namely, practice effects, differential carryover effects, and the potential for violations of certain statistical assumptions" (Keppel & Zedeck, 1989, p. 264). Practice effects occur when the subjects change systematically during the course of the experiment. Such changes may involve either a positive or a negative practice effect. A positive practice effect may show up as a result of an improvement, on the part of the subject, on the task that has been measured. On the other hand, a negative practice effect may show up due to fatigue or boredom. If fatigue were causing the change, lengthening the rest period between successive tasks may eliminate or minimize this problem. In the case where boredom is causing the change, monetary incentives may be used to keep the subjects motivated through the course of the experiment.

But as Keppel (1991) noted, "In most cases, however, researchers generally assume that practice effects will be present and that they can not be eliminated completely" (p. 335). A common solution to this problem is to introduce counterbalancing. Counterbalancing is a way of ordering treatments so that each treatment is administered an equal number of times first, second, third, and so on, in

particular sequences of conditions given to different subjects (Keppel & Zedeck, 1989). When using counterbalancing, two situations may arise: (a) there is an even number of levels of the treatment conditions; or (b) there is an odd number of levels of the treatment conditions.

When the number of levels of the treatment conditions, k, is an even number and the number of subjects, n, is some multiple of it, Girden (1992) provided the following guideline.

1, 2, k, 3, k-1, 4, k-2, etc.

For example, if there were two levels of treatments (k=2) and two subjects (n=2), the order of presentation would look schematically like Table 1. That is, subject one would be administered treatment A followed by treatment B. Subject two, however, would be administered treatment B followed by treatment A. In the case of four levels of treatment and four subjects, the order of presentation would look schematically like Table 2. That is, subject one would be administered the treatments in the following order. Treatment A would be first, treatment B would be second, treatment D would be third, and treatment C would be fourth. The order of presentation for subject two would be the following. Treatment B first, treatment C second, treatment A third, and treatment D fourth. The order of presentation for subjects three and four may be interpreted similarly from Table 2.

Insert Tables 1 and 2 About Here

When there are more subjects than levels of treatment, some of the orders of presentation will be repeated. For example, suppose there were eight subjects and four levels of

presentations in a particular study. Then, the first and the fifth subject would be administered the same first order of presentation. Similarly, the second and sixth subjects would be administered the same second order of presentation, etc. This situation is presented schematically in Table 3.

Insert Table 3 About Here

Thus, it follows that each treatment precedes each of the other treatments exactly once. That is, A precedes each of B, C, and D exactly once; B precedes each of A, C, and D exactly once; and, D precedes each of A, B, and C exactly once. In other words, each subject is given each treatment once and each treatment appears once in each level. This procedure helps to eliminate "the confounding that is surely present when only one sequence is used by counterbalancing the effect of practice over the treatment conditions equally" (Keppel & Saufley, 1980, p. 192). Continuing with the example of four levels of treatment, the first order of presentation would be 1, 2, 4, 3. The second order of presentation is derived by adding 1 to each of the numbers of the preceding order: 2 (1+1), 3 (2 +1), 1 (4+1 does not apply), 4 (3+1). This procedure would be continued until all the orders of presentation have been completed. Table 2 presents the completed order of presentation of the levels for the example with four levels and four treatment conditions.

When there is an odd number of levels of the treatment conditions, the first order of presentation is derived just as before. However, reversing the order of the first order and then repeating the procedure derives the remaining orders of presentation. For example, if five levels of treatment were to be administered, the first order of presentation

would be 1, 2, 5, 3, 4. As mentioned before, the second order of presentation would be 4, 3, 5, 2, 1 (i.e., simply reverse the first order of presentation). The third order of presentation would be derived, again, by adding 1 to each number in the preceding order. Thus, the third order of presentation would be 5 (4+1), 4 (3+1), 1 (5+1), 3 (2+1), 2 (1+1). Table 4 presents the completed order of presentation of the levels for five levels with five treatment conditions.

Insert Table 4 About Here

As stated by Maxwell and Delaney (1990), "Differential carryover occurs when the carryover effect of treatment condition 1 onto treatment condition 2 is different from the carryover effect of treatment condition 2 onto treatment condition 1" (p. 482). A common solution to this problem is to provide sufficient time between treatments so that the preceding treatment condition may dissipate completely from the subject's system. Unfortunately, unlike practice effects, differential carryover effects cannot be neutralized with counterbalancing (Keppel & Zedeck, 1989).

Assumptions for Repeated Measures Designs

Single-case repeated measures designs have the following three assumptions, as outlined by Stevens (1996): (a) independence of the observations; (b) multivariate normality; and (c) sphericity (sometimes called circularity). Of the three assumptions, sphericity is not necessary when the data are analyzed using the multivariate approach. However, just as violating the assumption for the observations is very serious for univariate analysis of variance (ANOVA) and for multivariate analysis of variance

(MANOVA), so it is here. Also, just as ANOVA and MANOVA are generally robust to violations of the multivariate normality, so that also applies here (Stevens, 1996).

The sphericity assumption is met when the variances of the differences of all treatment combinations are equal (i.e., the variance of the differences of treatments A and B equals the variance of the differences of treatments B and C, and so on). The variance of differences between two treatments is defined by

$$\sigma^2_{A-B} = \sigma^2_A + \sigma^2_B - 2\sigma_{AB}$$

where σ^2_A is the variance of a set of scores under treatment A, σ^2_B is the variance of another set of scores under treatment B, and σ_{AB} is the covariance of the two sets of scores. To illustrate this concept, suppose the set of scores in Table 5 have been obtained.

Insert Table 5 About Here

Once all the variances and covariances have been calculated, such values may be used to compute the variances of the differences of all treatments. These latter values may be used to determine if the assumption of sphericity has been met. To do so, the definition for the variance of the differences is applied to such values. For example, the variance of the difference between treatment A and treatment B would be

$$\sigma^2_{A-B} = \sigma^2_A + \sigma^2_B - 2\sigma_{AB}.$$

Substituting the values for σ^2_A, σ^2_B, and σ_{AB},

$$\sigma^2_{A-B} = 2.917 + 2.917 - 2(-2.08333) = 10.$$

However, since all the variances and covariances of the different treatment levels are the same, as reported in Table 6, all the variances of the differences here would be equal to 10. Thus, for this data set, the sphericity assumption is met.

Insert Table 6 About Here

A second approach to test the sphericity of a data set is to examine the matrix of orthonormal contrasts (Girden, 1992; Stevens, 1996). That is, sphericity is met if

$$C^T \Sigma C = \sigma^2 I$$

is true. Here, C is a matrix of (k-1) orthogonal contrasts, C^T is the transpose of C, Σ is the variance-covariance matrix and I is an identity matrix with σ^2 on the main diagonal and zeros elsewhere.

The first step in determining if the assumption of sphericity is met, is to create a set of (k-1) orthogonal contrasts. For the hypothetical data set in Table 5, a set of orthogonal contrasts is presented in Table 7. Contrast one compares the means for Treatments A and B (i.e., are there any differences between these two means?). Contrast two compares the combined means of treatments A and B with the mean of treatment C. Finally, contrast three compares the combined means of treatments A, B, and C with the mean of treatment D. Since the contrasts have means of zero and the sum of the cross-products of any two contrasts is zero, the contrasts are said to be orthogonal.

Insert Table 7 About Here

To construct matrix C in the above-mentioned formula, the contrasts are first normalized by multiplying each coefficient of a contrast by a value so that the sum of the

squared transformed coefficients equals one. This is accomplished by first squaring each coefficient in the contrast, then summing over the new squared coefficients, and finally, dividing each coefficient by the square root of the result. For example, squaring the coefficients of contrast one and summing over the new squared coefficients, $(1)^2 + (-1)^2 = 2$. Then, each coefficient in contrast one would be normalized by dividing each coefficient by the square root of 2. Similarly, squaring each coefficient in contrast three and summing over the new squared coefficients, $(1)^2 + (1)^2 + (1)^2 + (-3)^2 = 12$. Thus, each coefficient in contrast three would be normalized by dividing each coefficient by the square root of 12. These new coefficients are the coefficients of matrix C. For simplicity, such coefficients are presented in decimal form in Table 8. The transpose of matrix C is found by interchanging the rows and columns of matrix C. Next,

$$C^T \Sigma C = \sigma^2 I$$

is computed to determine if the sphericity assumption is met (Girden, 1992).

Insert Table 8 About Here

A third, "more direct way of determining variance of the difference is to calculate the difference between scores of two treatment levels (e.g. A-B) and determine the variance of these differences" (Girden, 1992, pp. 16-17). Using the data set in Table 5, the variance of the difference between treatment A and treatment B is 10. Similarly, all other variances of differences between any two treatments would be calculated. Such variances are assumed to be equal.

Correcting Violations to the Sphericity Assumption

When the sphericity assumption is not met, the actual level of statistical significance of the conventional (unadjusted) univariate *F* test on the repeated measures factor will exceed the nominal level (Barcikowski & Robey, 1984). That is, the Type I error rate will no longer be the preset α but a larger value. For example, instead of rejecting at the 0.05 level, perhaps the null hypothesis is being rejected at the 0.10 level. A common solution to this problem is to adjust the degrees of freedom by the correction factor epsilon (Girden 1992; Huynh & Feldt, 1976; Stevens, 1996). As O'Brien and Kaiser (1985) explained, "Epsilon measures nonsphericity: If epsilon equals one in the population, then sphericity holds and the traditional sampling distribution is designated. Reductions in epsilon indicate increasing degrees of nonsphericity and bring about suitable increases to the critical values for *F*" (p. 319).

Geisser and Greenhouse (1958) showed that the value of the epsilon is greater than or equal to 1/(k-1), where k is the number of treatments in the design. Geisser and Greenhouse also suggested evaluating the *F*-ratio at 1 and (n-1) degrees of freedom instead of evaluating the *F*-ratio at (k-1) and (k-1)(n-1) degrees of freedom. As Stevens (1996) explained, "Doing this makes the test very conservative, since adjustment is made for the worst possible case, and we don't recommend it" (p. 460). This procedure is conservative because smaller degrees of freedom correspond to a larger critical *F* value.

Another practical method for estimating the epsilon is epsilon hat. Such epsilon hat adjustment, although usually less severe than the Geisser and Greenhouse adjustment, is extremely tedious if done by hand. Maxwell and Delaney (1990) suggest using the following formula for computing epsilon hat:

$$\hat{e} = \frac{a^2\left(\overline{E}_{jj} - \overline{E}_{..}\right)^2}{(a-1)\left(\left(\sum\sum E_{jk}^2\right) - \left(2a\sum E_{j.}^2\right) + \left(a^2 E_{..}^2\right)\right)}$$

where

E_{jk} = an entry in the jth row and kth column of the sample covariance matrix,

E_{jj} = mean of variances along the diagonal in the same covariance matrix,

$E_{j.}$ = mean of the entries in the jth row of the same covariance matrix,

$E_{..}$ = mean of all entries in the same covariance matrix,

a = number of treatments.

However, since SAS and SPSS-X calculate this epsilon hat, the researcher need not be concerned with the computation's complexity. The researcher may, however, want to conceptually understand the theory behind the formula. Once epsilon hat is calculated, the value may be used to calculate yet another estimator for epsilon. This estimator, epsilon tilde, was introduced by Huynh and Feldt in 1976. They suggest using the following formula for computing epsilon tilde:

$$\tilde{e} = \frac{n(k-1)\hat{e} - 2}{(k-1)(n-1-(k-1)\hat{e})}$$

where

n = number of subjects in the study,

k = number of treatment levels,

e = defined as above.

Again, since SAS and SPSS-X calculate this epsilon tilde, the researcher may choose to concentrate on conceptually understanding the formula rather than concentrating on the calculations. Moreover, it can be shown that for any given n and k, epsilon tilde is greater than or equal to epsilon hat with the equality holding when epsilon hat equals 1/(k-1) (Huyhn & Feldt, 1976). Also, epsilon hat tends to underestimate epsilon, while epsilon tilde tends to overestimate epsilon. Consequently, the critical *F* value for epsilon tilde will typically be smaller than the critical *F* value for epsilon hat, thus leading to more rejections of the null hypothesis (Maxwell & Delaney, 1990).

Since the different epsilons will all be estimated by using a computer package, the researcher may choose to concentrate on deciding which epsilon to use. To do so, the researcher may follow the guidelines provided by Girden (1992, p. 21). These guidelines are:

1. If *e* is greater than .75, adjust the degrees of freedom by *e*;
2. If *e* is less than .75, adjust the degrees of freedom by the more conservative *e*; and
3. If nothing is known about *e*, adjust the degrees of freedom by the conservative *e*.

Using SPSS on the data in Table 4, the Geisser and Greenhouse epsilon, epsilon hat, was found to be 0.603. Similarly, the Huynh and Feldt epsilon, epsilon tilde, and the lower bound epsilon were calculated to be 1.00 and 0.333, respectively. Thus, following Girden's guidelines, the Geisser and Greenhouse epsilon, epsilon hat, would be used to adjust the degrees of freedom. Therefore, the adjusted degrees of freedom would be obtained by multiplying the original degrees of freedom by 0.603.

Example

The remainder of the paper will show how to do a single-case repeated measures analysis of variance. In doing so, the univariate as well as the multivariate approaches will be illustrated by means of the following example: Suppose a high school Algebra I teacher is interested in the effect of practice on the ability to solve algebra problems. First, four subjects, students, are administered an algebra test. Their scores are recorded as the number of problems solved correctly out of 20 problems. Then they are provided with practice on solving algebra problems. Finally, they are observed at posttest. However, if the teacher wanted to know whether the effects of practice persisted, the subjects could be tested again after three days and again one week following the practice session. The scores for this example are presented in Table 9.

Univariate Repeated Measures Analysis for Practice Effects Data

Analysis of variance (ANOVA) begins with the partitioning of the total variability in the experiment into two separate components, between treatments variability and within treatments variability. While this procedure is the same whether the ANOVA is for independent measures or for repeated measures designs, the two designs differ in the components of the between treatments variability. Figures 1 and 2 present the partitioning of the total variation for independent measures and for repeated measures, respectively.

Insert Figures 1 and 2 About Here

Notice that the independent measures design contains three sources of variability that contribute to between treatments variability: treatment effects, individual differences, and experimental error. On the other hand, because repeated measures designs use the

same subjects in every treatment, the variability between the treatments cannot be due to individual differences. Thus, there are only two sources of variability that contribute to between treatments variability: treatment effects and experimental error. It will therefore be this between treatments variability that will be used as the numerator of the *F*-ratio on subsequent calculations.

Because the subjects may come into the experiment with different levels of knowledge about algebra, these initial differences between the subjects may account for the variability within treatments. Another source of variability that contributes to within treatments variability is experimental error. This source of variability is introduced every time the researcher makes a measurement of the dependent variable. Notice, however, that there is no treatments effect contribution to within treatments variability since the set of scores are within the same treatment. For example, the four scores within the posttest may vary from each other but not because of treatment effect. Instead, such scores vary due to the individual differences and experimental error within that particular test.

Computing the Sums of Squares

The first sum of squares to be computed will be the total sum of squares variability (SOS_{total}). Once computed, this SOS_{total} will be partitioned into $SOS_{between\ treatments}$ and $SOS_{within\ treatments}$. Thus, symbolically,

$$SOS_{total} = SOS_{between\ treatments} + SOS_{within\ treatments}$$

However, because the subjects in repeated measures are being measured repeatedly, the variability due to individual differences ($SOS_{between\ subjects}$) needs to be measured. Consequently, the within treatments variability must be partitioned into individual differences and experimental error. Thus, symbolically,

$$SOS_{within\ treatments} = SOS_{between\ subjects} + SOS_{error}$$

Consequently,

$$SOS_{total} = SOS_{between\ treatments} + SOS_{between\ subjects} + SOS_{error}$$

The total variability in the experiment is found by

$$SOS_{total} = \sum X^2 - \frac{G^2}{N}$$

Here ΣX^2 is the sum of all squared scores, G is the sum of all the scores, and N is the number of scores in the entire experiment. Thus, for the data in Table 9,

$$SOS_{total} = 2576 - \frac{180^2}{16}$$

$$= 2576 - 2025$$
$$= 551$$

Next, the between treatment variability will be found using

$$SOS_{between\ treatments} = \sum \frac{T^2}{n} - \frac{G^2}{N}$$

Here T is the test total for each particular test, n is the number of subjects tested in each test administration, and G and N are defined as before. Thus, using the data in Table 9,

$$SOS_{between\ treatments} = \frac{11}{4} + \ldots + \frac{71^2}{4} - \frac{180^2}{16}$$

$$= 2541 - 2025$$
$$= 516$$

The within treatments variability is determined by adding the variability that is due to individual differences ($SOS_{between\ subjects}$) and the experimental error SOS_{error}. Thus,

$$SOS_{within\ treatmens} = SOS_{between\ subjects} + SOS_{error}$$

But,

$$SOS_{between\ subjects} = \Sigma n(x_{subject} - x_{grand})^2$$

Therefore, using the data in Table 9,

$$SOS_{between\ subjects} = 4(12 - 11.25 + \ldots + 11.25 - 11.25)^2$$

$$= 6.5$$

Once the SOS_{total}, $SOS_{between\ treatments}$, and $SOS_{between\ subjects}$ have been calculated, SOS_{error} may be found by using

$$SOS_{total} = SOS_{between\ treatments} + SOS_{between\ subjects} + SOS_{error}$$

and solving for SOS_{error}.

Thus,

$$SOS_{error} = 551 - 516 - 6.5$$

$$= 28.5$$

Partitioning the Degrees of Freedom

Like the total variability, the total degrees of freedom (*df*) may be partitioned into two components: between treatments *df* and within treatments *df*. However, just as in the case of within treatments variability, within treatments *df* need to be partitioned into between subjects *df* and experimental error *df*. Thus, symbolically,

$$df_{total} = df_{between\ treatments} + df_{within\ treatments}$$

$$= df_{between\ treatments} + df_{between\ subjects} + df_{error}$$

But,

$$df_{total} = N - 1$$

Thus, for the data in Table 9,

$$df_{total} = 16 - 1$$

$$= 15$$

Similarly,

$$df_{within\ treatments} = N - k$$

$$= 16 - 4$$

$$= 12$$

$$Df_{between\ subjects} = n - 1$$

$$= 4 - 1$$

$$= 3$$

But, using

$$df_{within\ treatments} = df_{between\ subjects} + df_{error}$$

$$df_{error} = df_{within\ treatments} - df_{between\ subjects}$$

Substituting,

$$df_{error} = 12 - 3$$

$$= 9$$

Lastly,

$$df_{between\ treatments} = k - 1$$

$$= 4 - 1$$

$$= 3$$

Computing the Mean Squares

The *F*-ratio is a ratio of two variances. Such variances, also called mean squares (*MS*), are computed by dividing a sum of squares by its corresponding *df*. The *MS* in the

numerator of the F-ratio is the between treatments MS. Such $MS_{between\ treatments}$ is found by applying the following formula:

$$MS_{between\ treatments} = \frac{SOS_{between\ treatments}}{df_{between\ treatments}}$$

The denominator of the F-ratio is MS_{error} and is found by applying the following formula:

$$MS_{error} = \frac{SOS_{error}}{df_{error}}$$

Therefore,

$$F = \frac{MS_{between\ treatments}}{MS_{error}}$$

A careful examination of the F-ratio reveals that

$$F = \frac{Treatment\ effect + error}{error}$$

Thus, when there is no treatment effect, the value of the F-ratio should be one. Conversely, if there is a treatment effect, the F-ratio will be larger than one.

For this example,

$$F = \frac{17.2}{3.17}$$

$$= 54.26$$

Table 10 presents the complete summary table for the univariate-repeated measures ANOVA. The critical value for F with 3 and 9 degrees of freedom is 3.86.

Notice that the calculated value of the F statistic, 54.26, is much larger than the expected value of one. Thus, there is a treatment effect. As a matter of fact, 93.65% of the total variation is due to the treatment, practice on solving algebra problems. Thus, the researcher may conclude that, on the average, the subjects benefited from the practice provided on solving algebra problems.

Insert Table 10 About Here

Multivariate Approach

A second solution to the problem arising when the sphericity assumption has been violated is to use multivariate analysis of variance (MANOVA) methods; as Girden (1982) noted, "sphericity is not an assumption here" (p. 23). However, MANOVA methods assume multivariate normality. Nonetheless, violations of the multivariate normality assumptions are generally regarded as less serious than violations of the sphericity assumption (Maxwell & Delaney, 1990).

The MANOVA analysis is done not on the original scores but on new latent/synthetic variables constructed from the measured variables. These new variables are obtained by subtracting adjacent repeated measures (e.g., Pretest - Posttest, Posttest - 3 Days After, 3 Days After - 1 Week After). These new variables are then used to compute the F statistic. Table 11 presents these new variables.

Insert Table 11 About Here

Once all the means, variances, and covariances have been calculated, Hotelling's T^2 may be used to compute the F statistic.

$$T^2 = nyX^{-1}y^T$$

where n is the number of subjects in the study, y is the row vector of means, X^{-1} is the inverse of the variance-covariance matrix, and y^T is the transpose of y (i.e. the column vector of means). Therefore, using the data in Tables 10 and 11

$$T = 4\begin{bmatrix}-7 & -5 & -2.5\end{bmatrix}\begin{bmatrix}10 & -1.67 & 2.33\\ -1.67 & 3.33 & -3.67\\ 2.33 & -3.67 & 4.3\end{bmatrix}^{-1}\begin{bmatrix}-7\\ -5\\ -2.5\end{bmatrix}$$

$$= 1052.12$$

However, T^2 may be converted to F by means of

$$F = \frac{n-k+1}{(n-1)(k-1)}T^2$$

with (k - 1) and (n - k + 1) degrees of freedom. Thus,

$$F = \frac{4-4+1}{(3)(3)}(1052.12)$$

$$=116.90$$

with 3 and 3 degrees of freedom. The critical value for F with 3 and 3 degrees of freedom is 9.28. Thus, again, the results are statistically significant, if that means anything to anyone. Consequently, the researcher may conclude that, on the average, the subjects benefited from the practice provided on solving algebra problems.

Should the Univariate or the Multivariate Approach be Used?

When the sphericity assumption has been violated, the actual level of statistical significance of the unadjusted univariate *F* test will no longer be the preset α. In other words, instead of committing a Type I error 5% of the times, such an error might be committed 10 or 15% of the time. Thus, the researcher faces two options: adjust the degrees of freedom or use the multivariate approach. However, even if the researcher opts for adjusting the degrees of freedom, the obtained results are only approximate. On the other hand, when the multivariate normality assumption has been met, the actual α level of the multivariate approach is guaranteed mathematically to be equal to the preset α level. Thus, when the researcher's concern is the probability of falsely rejecting the null hypothesis, the multivariate approach is suggested (Maxwell & Delaney, 1990).

When the sphericity criterion holds, the univariate test is more powerful than the multivariate test (Maxwell & Delaney, 1990). However, "sphericity almost always is violated"(Girden, 1992, p. 26). When sphericity has been violated, neither test exceeds the other in terms of power. However, "for moderate sample sizes, the multivariate test ranges from somewhat less powerful to much more powerful than the mixed-model test"(Maxwell & Delaney, 1990, p. 605). Thus, it follows that when n, the number of subjects in the study, exceeds k, the number of levels of the repeated factor, by a few, the multivariate test is more powerful than the univariate test.

This paper presented how to analyze repeated measures designs using the univariate as well as the multivariate approach. Each design has its own assumptions to meet. However, the sphericity assumption of the univariate approach is almost always violated. On the other hand, even when the normality assumption of the multivariate

approach is violated, such violations are generally regarded as less serious than violations of the sphericity assumption. Therefore, when the researcher's concern is committing a Type I or a Type II error and several assumptions hold, the multivariate approach is suggested.

References

Barcikowski, R.S., & Robey, R.R. (1984). Decisions in single group repeated measures analysis: Statistical test and three computer packages. The American Statistician, 38, 148-150.

Girden, E.R. (1992). ANOVA: Repeated measures. Newbury Park, CA: Sage.

Geisser, S., & Greenhouse, S.W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics, 29, 885-891.

Huynh, H. & Feldt, L.S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. Journal of Educational Statistics, 1, 69-82

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Keppel, G. & Saufley, W.H. (1980). Introduction to design and analysis: A student's handbook. San Francisco, CA: W.H. Freeman and Company.

Keppel, G. & Zedeck, S. (1989). Data analysis for research designs. New York: W.H. Freeman and Company.

Maxwell, E.S. & Delaney, H.D. (1990). Designing experiments and analyzing data. Belmont, CA: Wadsworth Phublishing Company.

Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (1996). Applied linear statistical models (4th ed.). Chicago, IL: Irwin.

O'Brien, R.G. & Kaiser, M.K. (1985). Manova methods for analyzing repeated measures designs: An extensive primer. Psychological Bulletin, 97, 316-333.

Stevens, J. (1996). Applied multivariate statistics for the social sciences. Mahwah, NJ: Lawrence Erlbaum.

Table 1

Example of counterbalancing for two subjects

	Order of Treatments	
Subject	1	2
1	A	B
2	B	A

Table 2

Example of Counterbalancing for four subjects

	Order of Treatments			
Subject	1	2	3	4
1	A	B	D	C
2	B	C	A	D
3	C	D	B	A
4	D	A	C	B

Table 3

Example of Counterbalancing for eight subjects

	Order of Treatments			
Subject	1	2	3	4
1	A	B	D	C
2	B	C	A	D
3	C	D	B	A
4	D	A	C	B
5	A	B	D	C
6	B	C	A	D
7	C	D	B	A
8	D	A	C	B

Table 4

Example of Counterbalancing for five subjects

	Order of Treatments				
Subject	1	2	3	4	5
1	1	2	5	3	4
2	4	3	5	2	1
3	5	4	1	3	2
4	1	5	2	4	3
5	2	1	3	5	4

Table 5

Hypothetical data set to Illustrate Sphericity

	Order of Treatments			
Subject	A	B	C	D
1	1	12	15	20
2	2	10	17	17
3	3	8	14	16
4	5	9	13	18

Table 6

Variance-Covariance Matrix

	Order of Treatments			
Subject	A	B	C	D
1	2.917	-2.08333	-2.08333	-2.08333
2	-2.08333	2.917	-2.08333	-2.08333
3	-2.08333	-2.08333	2.917	-2.08333
4	-2.08333	-2.08333	-2.08333	2.917

Table 7

Orthogonal Contrasts for Table 3 data

	Contrasts		
Treatment	C_1	C_2	C_3
A	1	1	1
B	-1	1	1
C	0	-2	1
D	0	0	-3

Table 8

Orthonormal Matrix C

	Contrasts		
Treatment	C_1	C_2	C_3
A	0.707	0.408	0.289
B	-0.707	0.408	0.289
C	0	-0.816	0.289
D	0	0	-0.866

Table 9

Number of Correct Problems out of 20

	Test Session			
Subject	Pretest	Posttest	3 Day After	1 Week After
1	1	12	15	20
2	2	10	17	17
3	3	8	14	16
4	5	9	13	18

Table 10

Summary of ANOVA for Repeated Measures on a Single Factor

Source	SS	df	MS	F	Eta^2
Subjects	6.5	3	2.17		
Treatments	516	3	172	54.26	
Residual	28.5	9	3.17		
Total	551	15			

Table 11

Differences Between Adjacent Repeated Measures

Subject	Pretest-Posttest	Posttest-3 Days After	3 Days After- 1 Week After
1	-11	-3	-5
2	-8	-7	0
3	-5	-6	-2
4	-4	-4	-3

Figure 1. Partitioning of variance for an independent measures design

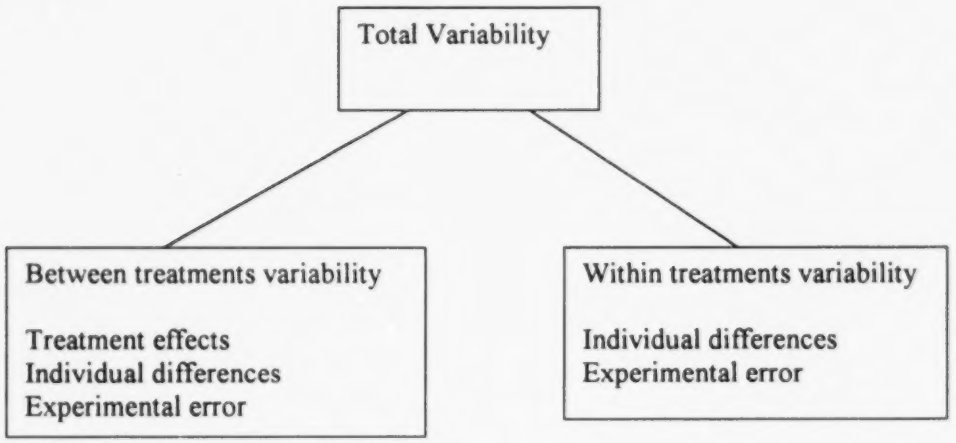

Figure 2. Partitioning of variance for a repeated measures design